AD_____

Award Number: DAMD17-97-1-7202

TITLE: Investigation of Genetic Algorithms for Computer-Aided Diagnosis

PRINCIPAL INVESTIGATOR: Matthew A. Kupinski

CONTRACTING ORGANIZATION: University of Chicago
Chicago, Illinois 60637

REPORT DATE: October 1999

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1999	Annual (1 Oct 98 – 30 Sep 99)

4. TITLE AND SUBTITLE
Investigation of Genetic Algorithms for Computer-Aided Diagnosis

5. FUNDING NUMBERS
DAMD17-97-1-7202

6. AUTHOR(S)
Matthew A. Kupinski

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Chicago
Chicago, Illinois 60637

E-MAIL:
m-kupinski@uchicago.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for release; Distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Computer-aided diagnosis has the potential of substantially increasing diagnostic accuracy in mammography. Using a computer to double-check a radiologist's findings is becoming more popular and more important as the public learns that the best defense against breast cancer is early detection. The University of Chicago is currently developing computerized schemes to detect cancers in digital mammograms. We use a pattern classification system known as an artificial neural network (ANN) to classify certain regions of the digital mammograms as cancerous or non-cancerous. ANNs are trained pattern classification devices that take, as inputs, features extracted from regions in the mammograms and output the classification. Previously, we reported on the use of genetic algorithms for feature selection. Currently, there are a total of 42 features extracted from the various regions in each mammogram. A subset of those 42 features must be chosen as inputs for the ANN. The goal of the past year's research was to investigate methods of feature selection and pattern classification in order to improve upon the overall performance of CAD systems.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
15

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

MK Where copyrighted material is quoted, permission has been obtained to use such material.

MK Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

MK Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Matthew A Krupath 11/17/99

PI - Signature Date

Table of contents

4 Introduction

Breast cancer is a major cause of death among women over the age of forty [1]. Mammography is the most effective diagnostic procedure for the early detection of breast cancer [2, 3]. Mammography is not, however, perfect. Between 10-30% of women who have breast cancer and undergo mammography have negative mammograms [4–7]. Of these, radiologists have determined, retrospectively, that two-thirds of the cancers could have been detected [5, 6, 8, 9]. One possible means by which to decrease this number is to have two radiologists read the mammograms. This method has been shown to increase sensitivity by as much as 15%, [10, 11] but can be costly both financially and with respect to time. A computer-aided diagnostic scheme may act as an inexpensive second reading method. The final decision would be made by the radiologist.

The proposed research seeks to answer questions that arise when using pattern classifiers in decision making applications. Problems occur when the number of inputs to the pattern classifier become large. For this reason, genetic algorithms and other *feature selection* techniques are currently being studied to alleviate this problem. The purpose of this proposed research is to study and develop feature selection and pattern classification methods to improve the performance of CAD schemes. Specific emphasis will be placed using the developed methods in the computerized detection of mass lesions in mammography.

5 Body

5.1 Investigation of Feature Selection

Feature selection is the task of selecting a useful and robust subset of features to be used within a classifier. To gain a better understanding of the difficulties associated with selecting features, we examined a relatively simple feature-selection problem using A_z (area under the ROC curve) as a performance measure. By studying this simple problem we hope to gain understanding of more complicated feature selection methods such as genetic algorithms. Let us consider the following ideal situation: We have a total of D independent features with the first r features having theoretical A_z values of $A_z^{(1)}$ and the remaining $D-r$ features having theoretical A_z values of $A_z^{(2)}$, where $A_z^{(1)} > A_z^{(2)}$. Because the features in this situation are independent, we conclude that the D random variables denoting the measured $\boldsymbol{A_z}$ values are also independent with density functions given by $p_{A_z}^{(1)}(A_z)$ for the r features with theoretical A_z values of $A_z^{(1)}$ and $p_{A_z}^{(2)}(A_z)$ for the $D-r$ features with theoretical A_z values of $A_z^{(2)}$. Similarly, the distribution functions are given by $P_{A_z}^{(1)}(A_z)$ for the r features with theoretical A_z values of $A_z^{(1)}$ and $P_{A_z}^{(2)}(A_z)$ for the $D-r$ features with theoretical A_z values of $A_z^{(2)}$.

The task in this situation is to select the d features (where $d \leq r$) that have the largest measured $\boldsymbol{A_z}$ values. Because, however, the measured $\boldsymbol{A_z}$ values have a distribution associated with them, there is a measurable probability that one or more of the "worse" features (those features with a theoretical A_z value of $A_z^{(2)} < A_z^{(1)}$) will be selected. Using order statistics [12–15], we have derived the probability that an optimal subset of features will be selected in the situation described above:

$$\frac{r!}{(d-1)!(r-d)!} \int_0^1 dA_z \; p_{A_z}^{(1)}(A_z)(1 - P_{A_z}^{(1)}(A_z))^{d-1} P_{A_z}^{(1)}(A_z)^{r-d} P_{A_z}^{(2)}(A_z)^{D-r}, \qquad (1)$$

where the integration is from 0 to 1 because A_z values are bound between 0 and 1. In theory, the probability in the situation where each independent feature has a different theoretical A_z value could be computed, but it is computationally impractical.

5.1.1 Results to Date

Figure 1(a) plots the probability of an optimal subset consisting of 4 features being selected as a function of the total number of features D (See Eqn. 1). In this plot the total number of features with theoretical $A_z = A_z^{(1)}$ was 4, $A_z^{(1)}$ was set at 0.70, and $A_z^{(2)}$ was fixed at 0.60. The dataset size s was also varied from 100 to 1000 where there were equal numbers of abnormal and normal observations, *i.e.*, $s_a = s_n = s/2$. As Fig. 1(a) shows, with small dataset sizes the probability of selecting an optimal subset of features drops quickly as the total number of features D increases. Figure 1(b) shows similar plots ($d = 4$, $r = 4$) but with higher theoretical A_z values, *i.e.*, $A_z^{(1)} = 0.8$ and $A_z^{(2)} = 0.7$. Comparison of Figs. 1(a) and 1(b) indicates that, although the differences in theoretical A_z values ($A_z^{(1)} - A_z^{(2)}$) are the same, the probabilities of selecting an optimal subset of features vary for identical dataset sizes. These findings indicate that the probabilities of selecting an optimal subset of features

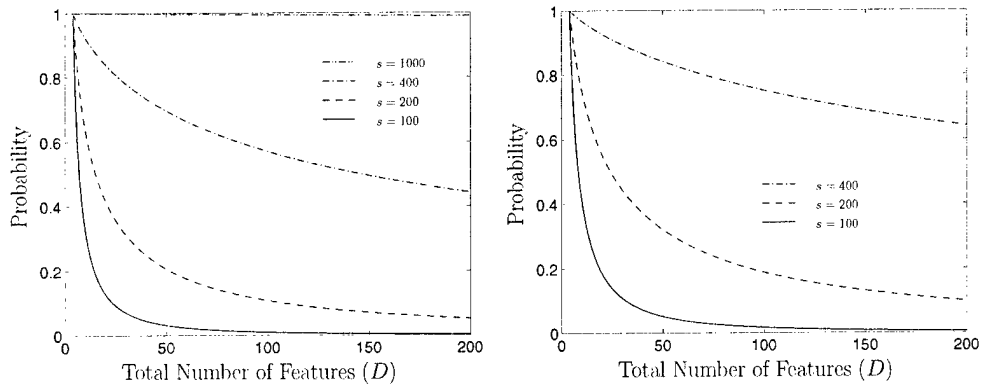


Fig 1: A plot of the probability of selecting an optimal subset consisting of $d = 4$ features from a total of D features. There are a total of $r = 4$ features with a theoretical A_z value of $A_z^{(1)} = 0.7$ for (a) and $A_z^{(1)} = 0.8$ for (b). There are also $D - r$ features with a theoretical A_z value of $A_z^{(2)} = 0.6$ for (a) and $A_z^{(2)} = 0.7$ for (b). The probability of selecting an optimal subset of features is also plotted for various dataset sizes s.

depend on the theoretical A_z values of the features and not solely on the differences between the theoretical A_z values of the "good" and "bad" features.

In a second study, we simulated D features using Gaussian distributions, where d features had theoretical A_z values of $A_z^{(1)} = 0.68$, and $D - d$ features had theoretical A_z values of $A_z^{(2)} = 0.60$. The d features with the highest measured $\boldsymbol{A_z}$ values were then combined using linear discriminant analysis to merge the d-dimensional features to a scalar decision variable. The $\boldsymbol{A_z}$ value of the classifier was measured using that decision variable data. The same dataset employed to select the d features was used to determine the parameters of the linear discriminant that merged the d features. We also tested the classifier on an independent dataset of 1000 samples. This process was repeated 100 times for each combination of parameters to obtain an average training dataset $\boldsymbol{A_z}$ and testing dataset $\boldsymbol{A_z}$ values for the classifier. Figure 2 shows a plot, for various total numbers of features D, of the average training and testing dataset $\boldsymbol{A_z}$ values as a function of the dataset size s. The thin solid line in Fig. 2 is the theoretical A_z value of 4 independent Gaussian features, with equal variances and individual A_z values of 0.68, merged using linear discriminants. The curves above the theoretical line are the average training dataset $\boldsymbol{A_z}$ values, and the curves below the theoretical line are the average testing dataset $\boldsymbol{A_z}$ values. Figure 2 indicates that bias is introduced when the same datasets are used to select and merge features. The bias is enhanced when the dataset size is small and there are a large number of features D; these are the same conditions under which selection of an optimal subset of features is the most difficult (see Figs. 1(a)-1(b)). Hence, a suboptimal subset of features is most likely selected and bias is introduced because we are employing the same dataset to both select features and determine the parameters of the classifier.

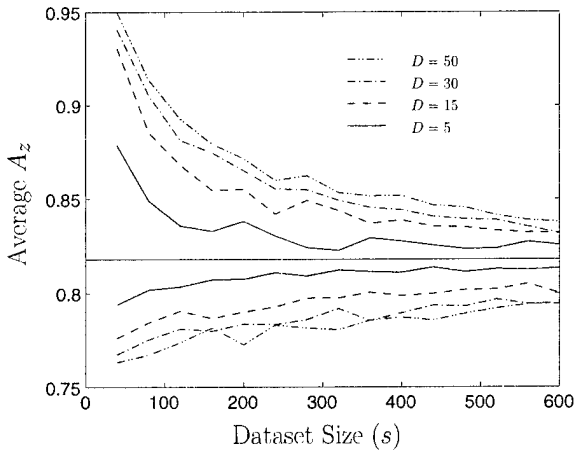


Fig 2: A total of $r = 4$ Gaussian features were simulated with theoretical A_z values of $A_z^{(1)} = 0.68$ and $D - r$ Gaussian features were simulated with theoretical A_z values of $A_z^{(2)} = 0.6$. Features were sampled 100 different times and the top $d = 4$ features were selected based on the measured $\boldsymbol{A_z}$ values of the individual features. The selected features were then merged using linear discriminants and the training dataset and testing dataset $\boldsymbol{A_z}$ values were computed. The thin solid line at an A_z of 0.818 is the theoretical true A_z if the 4 independent Gaussian ($A_z = 0.68$) features are merged using linear discriminants. The curves above this theoretical line are the average training dataset $\boldsymbol{A_z}$ values and the curves below the theoretical line are the average testing dataset $\boldsymbol{A_z}$ values. The same dataset used to select the features was employed in determining the parameters of the linear discriminants. A substantial amount of bias is introduced for small dataset sizes s and a large total number of features D.

5.2 Investigation of Bayesian ANNs

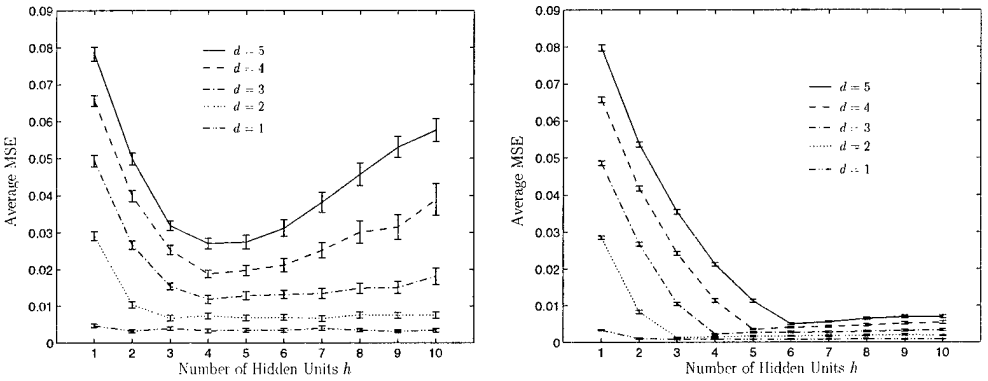
In order to perform feature selection with artificial neural networks (ANNs), one must have a performance or fitness measure in which to optimize. If one were to fully train an ANN and then test the ANN's performance on the training dataset, the results will often be artificially high, *i.e.*, the ANN overtrained. One method of circumventing this problem is the use round-robin methodology [16] in which numerous ANNs are trained with different subsets of the data and then tested on the parts of the data left out. This method has been shown to work well but is time consuming. We have studied the use of Bayesian ANNs for classification purposes. We have found that Bayesian ANNs are more accurate and quicker to train than conventional ANNs using round-robin methodology. Bayesian ANNs regularize training in a much different manner, *i.e.*, they use a prior term in the cost function to penalize complicated (or overtrained) ANN solutions. This allows for more rapid ANN training without overtraining and is, thus, more practical as a pattern classifier and for feature selection tasks.

5.2.1 Results to Date

Figure 3(a) shows the performance of the Bayesian ANN with varying numbers of hidden units h and input dimensions d for a fixed signal-to-noise ratio $SNR = 1.26$ and dataset size $s = 200$. For lower dimensions ($d = 1$, $d = 2$, and $d = 3$), the average mean squared error (MSE) between the optimal decision variable and the Bayesian ANN approximation of that decision variable decreases as the number of hidden units increases and then becomes relatively constant after a certain threshold. For $d = 2$, the MSE becomes relatively constant after 3 hidden units, while for $d = 3$, the MSE flattens out after 4 hidden units. More parameters are required to better approximate the optimal mapping function as the number of dimensions increases. For $d = 4$ and $d = 5$, there is a pronounced minimum in the MSE. These results indicate that as the number of dimensions increases, the Bayesian ANN does not have enough data to approximate properly the optimal mapping function. Consequently, a tradeoff exists between simpler solutions (*i.e.*, few parameters $\vec{w}$) that cannot match the ideal observer due to under-parameterization and more complex solutions with numerous parameters $\vec{w}$ that cannot be properly determined due to the lack of data.

5.3 Investigation of RGI Filtering

In CAD, the performance of a pattern classifier is limited by the performance of the initial detection filtering. For example, if the pattern classifier operates at a specificity of 90% but the initial detection algorithm returns 50 false detections per image, then the final performance will be 5 false positives per image which is unacceptable for clinical implementation. We have previously analyzed the use of a constraint function and the radial gradient index (RGI) feature in the segmentation of mass lesions in mammograms [17]. In this work, we will extend the use of RGI feature to a non-linear filtering method which can be used in the initial detection phase of a mass detection scheme. Comparisons between this new filtering method and previous methods will be presented.



(a) (b)

Fig 3: The effect of the number of hidden units h on the accuracy of Bayesian ANNs with a fixed $SNR = 1.26$ and dataset size (a) $s = 200$ and (b) $s = 1000$. With a limited training dataset (a), the Bayesian ANN cannot properly approximate the optimal mapping function at higher dimensions ($d = 4$ and $d = 5$) but does not have a problem with a larger training dataset (b). The error bars represent $\pm\frac{1}{2}\sigma$.

5.3.1 Results to Date

In order to evaluate the overall performance of this filtering technique, we employed a database of 112 mammograms containing 64 malignant mass lesions (118 visible lesions) digitized on a Lumisys 100 digitizer using a 100 μm pixel size and 12-bit gray-level quantization. The images were subsampled to an effective pixel size of 0.5 mm. Each visible lesion was outlined by a radiologist experienced in mammography. For each image, an RGI filtered image was generated using a skip factor of 4. These images were then thresholded at many different RGI threshold values ranging from -1 to 1 and the centers of those regions passing both the RGI threshold and the various size cutoffs were compared with the radiologist's outlines for each image. If the center of a region was contained within the radiologist's outline for that image, then that lesion was considered to be detected. Figure 4 shows the FROC curves for the RGI filtering technique at various size cutoffs and where the implicit FROC decision variable is the RGI threshold value. Each FROC curves exhibits the interesting property of beginning and ending at similar locations near $(0, 0)$ in FROC space. At a very high RGI threshold, no pixels will pass the threshold and there will be no true detections and also no false detections. At a very low threshold, however, there will only be one connected region returned because every point within the image passes the threshold so the sensitivity and the number of false detections per image will be low.

One can also see in Fig. 4 that increasing the size cutoff causes the FROC curve to shift to the left and down. It is important to note that for a size cutoff of 1 (regions must be greater than 1 pixel), the shift is much greater to the left than it is down from a size cutoff of 0 pixels (no minimum size). The false-positive rate is substantially reduced at a minimal cost to the sensitivity of the method. This agrees with the assumption that many of the detections of size 1 pixel are due to random fluctuations and not actual abnormalities in the

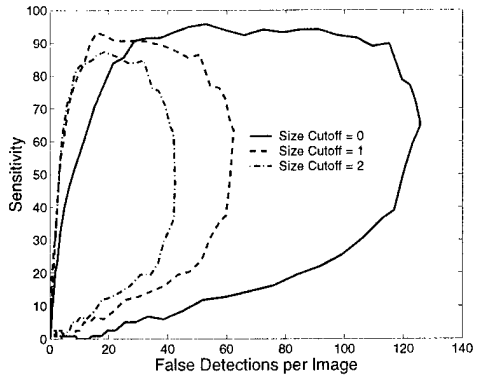


Fig 4: RGI filtering FROC curves for various minimum size cutoffs using the RGI threshold as the decision variable. Note that there is a large decrease in the false-detection rate when the minimum size cutoff in increased from 0 to 1 without a large decrease in the sensitivity.

image.

The previous method employed during the initial detection phase of the mass detection method was the bilateral subtraction technique [18]. On the same database, bilateral subtraction yielded a sensitivity of 64% at 48 false detections per image. The RGI filtering technique, as can be seen in Fig. 4, can yield a sensitivity of 93% at 16 false detections per image which represents a substantial improvement. This, however, is just the first step in the computerized detection scheme. In order to evaluate the performance of the overall technique, we implemented a simple pattern classifier and applied it to the regions returned by RGI filtering and bilateral subtraction. Each point returned by RGI filtering was used as a seed point for our lesion segmentation algorithm described in [17]. The lesion segmentation algorithm returns a contour which "best" delineates the potential lesion. Using this information, along with the image function $f(x, y)$, we extracted three features; the RGI, the contrast [19], and the average gradient strength along the segmented contour. Linear discriminant analysis [20, 21], trained on an independent dataset, was used to distinguish between actual lesions and false detections. The training datasets for both initial detection methods consisted of the true lesions detected by each method and a randomly chosen subset of false detection returned by each method. FROC curves showing the performance of this combined scheme are shown in Fig. 5 for both the previous method of bilateral subtraction and RGI filtering. The sensitivity plotted in Fig. 5 is the by-patient sensitivity. Note that for the RGI filtering curve, the RGI threshold was fixed at 0.74 and the size cutoff was fixed at 1 which corresponds to the $(93\%, 16)$ point in FROC space in Fig. 4, and the linear discriminant threshold value was employed to sweep out the FROC curves. The same features were used in both the RGI filtering FROC curve and the bilateral subtraction curve.

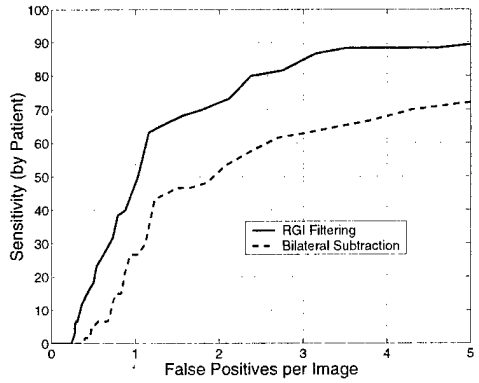


Fig 5: FROC curves for the RGI filtering technique and bilateral subtraction using a simple pattern classify to reduce the number of false detections returned by both methods. A total of three features were used in the pattern classifier for both methods.

6 Conclusions

We have studied some of the fundamental properties of feature selection. We have found that the probability of selecting an optimal subset of features rapidly decreases as the sample size decreases and the total number of features from which to select increases. Understanding the limitation of feature selection will help us select (using methods such as 1D analysis and genetic algorithms) a useful and robust subset of features to be used in the computerized detection of mass lesions in mammography.

We have also studied the use of Bayesian artificial neural networks in classification tasks. We have found that Bayesian ANNs produce more accurate and, yet, robust solutions to classification problems. Bayesian ANNs also train more rapidly than do conventional ANNs using round-robin methodology. This information will be used design more accurate and robust pattern classifiers for the computerized detection of mass lesions in mammography.

We have introduced a new initial filtering scheme to detect mass lesions in mammography. The performance of feature selection methods and of pattern classifiers is limited by the performance of the initial detection algorithm. We have shown that RGI filtering substantially outperformed the previous method of detecting mass lesions known as bilateral subtraction.

In the future we will use the RGI filtering technique to detect suspicious image regions in mammograms. We will then extract features from these regions and select a useful subset of features taking the knowledge we have gained from our feature selection studies into account. Finally, these features will be classified using a Bayesian artificial neural network.

7 Papers and Presentations

The following papers have been accepted or submitted to peer review journals:

- "Feature Selection with Limited Datasets." Matthew A. Kupinski and Maryellen L. Giger. *Medical Physics*, Vol. 26, Iss. 10, Pgs. 2176-2182, 1999. [22]
- "Multiobjective Genetic Optimization of Diagnostic Classifiers with Implications for Generating ROC Curves." Matthew A. Kupinski and Mark A. Anastasio. *IEEE Transactions on Medical Imaging*, Vol. 18, Iss. 8, Pgs. 675-685, 1999. [23]
- "Optimization and FROC Analysis of Rule-Based Detection Schemes Using a Multiobjective Approach." Mark A. Anastasio, Matthew A. Kupinski and Robert M. Nishikawa. *IEEE Transactions on Medical Imaging*, Vol. 17, Iss. 6, Pgs. 1089-1093, 1998. [24]
- "Ideal Observer Estimation Using Bayesian Classification Neural Networks." submitted to *IEEE Transactions on Medical Imaging.*
- "Computerized Detection of Mass Lesions in Mammography Based on Radial Gradient Index." submitted to *Medical Physics.*

The following presentations have been given:

- "Feature Selection with Limited Datasets," Matthew A. Kupinski and Maryellen L. Giger, RSNA 1998.
- "Multiobjective Optimization of Diagnostic Classifiers and its Relationship to ROC Analysis and the Ideal Observer," Mark A. Anastasio and Matthew A. Kupinski, Future Directions in Nuclear Medicine Physics and Engineering, Chicago, IL, 1999.
- "Multiobjective Optimization of Diagnostic Classifiers: Pareto Optimality and the Ideal Observer," Mark A. Anastasio and Matthew A. Kupinski, Eighth Far West Image Perception Conference, Alberta, Canada.
- "Ideal Observer Estimation With Bayesian Classification Neural Networks," Matthew A. Kupinski, Darrin C. Edwards, and Maryellen L. Giger, Eighth Far West Image Perception Conference, Alberta, Canada.
- "Bayesian Artificial Neural Networks in the Computerized Detection of Mass Lesions," Matthew A. Kupinski, Darrin C. Edwards, Maryellen L. Giger, and Alexandra E. Baehr, American Association of Physicists in Medicine, Nashville, Tennessee, 1999.
- "Computerized Detection of Mass Lesions in Digital Mammography Using Radial Gradient Index Filtering," Matthew A. Kupinski and Maryellen L. Giger, to be presented at RSNA 1999.

References

[1] E. Silverberg, C. C. Boring, and T. S. Squires, *Cancer Statistics*, vol. 40. 1990.

[2] L. W. Basset and R. H. Gold, *Breast Cancer Detection. Mammography and Other Methods in Breast Imaging.* Grune and Stratton, 1987.

[3] J. Lissner, M. Kessler, and G. Anhalt, "Developments in methods for early detection of breast cancer," in *Early Breast Cancer* (J. Aandler and J. Baltzer, eds.), (Berlin), Springer-Verlag, 1984.

[4] I. Andersson, "What can we learn from interval carcinomas?," *Recent Results in Cancer Research*, vol. 90, pp. 191–193, 1984.

[5] C. J. Baines, A. B. Miller, and C. Wall, "Sensitivity and specificity of first screen mammography in the canadian national breast screening study. a preliminary report from five centers," *Radiology*, vol. 160, pp. 295–298, 1986.

[6] J. E. Martin, M. Moskowitz, and J. R. Milbrath, "Breast cancers midded by mammography," *American Journal of Roentgenology*, vol. 132, pp. 737–739, 1979.

[7] S. R. Pollei, F. A. Mettler, S. A. Bartow, G. Moradian, and M. Moskowitz, "Occult breast cancer. prevalence and radiographic detectability," *Radiology*, vol. 16, pp. 459–462, 1987.

[8] J. B. Buchanan, J. S. Spratt, and L. S. Heuser, "Tumor growth, doubling times, and the inability of the radiologist to diagnose certain cancers," *Radiologic Clinics of North America*, vol. 21, pp. 115–126, 1983.

[9] T. Holland, M. Mrvunac, J. H. C. L. Hendrirks, and B. Bekker, "So-called interval cancers of the breast. pathologic and radiographic analysis," *Cancer*, vol. 49, pp. 2527–2533, 1982.

[10] E. L. Thurfjell, K. A. Lernevall, and A. A. Taube, "Benefit of independent double reading in a population-based mammography screening program," *Radiology*, vol. 191, pp. 241–244, 1994.

[11] C. E. Metz and J.-H. Shen, "Gains in accuracy from replicated readings of diagnostic images: Prediction and assessment in terms of ROC analysis," *Medical Decision Making*, vol. 12, pp. 60–75, 1992.

[12] H. A. David, *Order Statistics.* Wiley Series in Probability and mathematical Statistics, New York: John Wiley & Sons, Inc., 1970.

[13] J. D. Gibbons, I. Olkin, and M. Sobel, *Selecting and Ordering Populations: A New Statistical Methodology.* Wiley Series in Probability and Mathematical Statistics, New York: John Wiley & Sons, 1977.

[14] J. I. Marden, *Analyzing and Modeling Rank Data.* Monographs on Statistics and Applied Probability, London: Chapman & Hall, 1995.

[15] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics.* Wiley Series in Probability and Mathematical Statistics, New York: John Wiley & Sons, Inc., 1992.

[16] M. Kupinski, M. L. Giger, P. Lu, and Z. Huo, "Computerized detection of mammographic lesions: Performance of artificial neural network with enhanced feature extraction," in *SPIE*, vol. 2434, pp. 598–605, 1995.

[17] M. A. Kupinski and M. L. Giger, "Automated seeded lesion segmentation on digital mammograms," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 510–517, 1998.

[18] F.-F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique," *Medical Physics*, vol. 21, pp. 445–452, 1994.

[19] B. Zheng, Y.-H. Chang, and D. Gur, "On the reporting of mass contrast in cad research," *Medical Physics*, vol. 23, no. 12, pp. 2007–2009, 1996.

[20] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition.* Applications of Mathematics, New York: Springer-Verlag Inc., 1996.

[21] K. Fukunaga, *Statistical Pattern Recognition.* San Diego: Academic Press, 1990.

[22] M. A. Kupinski and M. L. Giger, "Feature selection with limited datasets," *Medical Physics*, vol. 26, pp. 2176–2182, 1999.

[23] M. A. Kupinski and M. A. Anastasio, "Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curves," *IEEE Transactions on Medical Imaging*, vol. 18, pp. 675–685, 1999.

[24] M. A. Anastasio, M. A. Kupinski, and R. M. Nishikawa, "Optimization and FROC analysis of rule-based detection schemes using a multiobjective approach," *IEEE Transactions on Medical Imaging*, vol. 17, pp. 1089–1093, 1998.